THEORY AND METHODS

Uncertainty analysis: an example of its application to estimating a survey proportion

Anne M Jurek, George Maldonado, Sander Greenland, Timothy R Church

....

See end of article for authors' affiliations

....

Correspondence to:
Dr A M Jurek, Department of Pediatrics, University of Minnesota, Mayo Mail Code 715, 420 Delaware St SE, Minneapolis, MN 55455, USA; jure0007@umn.edu

Accepted 16 February 2007
....

Uncertainty analysis is a method, established in engineering and policy analysis but relatively new to epidemiology, for the quantitative assessment of biases in the results of epidemiological studies. Each uncertainty analysis is situation specific, but usually involves four main steps: (1) specify the target parameter of interest and an equation for its estimator; (2) specify the equation for random and bias effects on the estimator; (3) specify prior probability distributions for the bias parameters; and (4) use Monte-Carlo or analytic techniques to propagate the uncertainty about the bias parameters through the equation, to obtain an approximate posterior probability distribution for the parameter of interest. A basic example is presented illustrating uncertainty analyses for four proportions estimated from a survey of the epidemiological literature.

Epidemiologists use methods for interpreting the impact of bias sources (systematic errors) on study results that range from ignoring biases, to evaluating biases qualitatively, to evaluating biases quantitatively. Uncertainty analysis is a family of methods, established in engineering and policy analysis[1 2] but relatively new to epidemiology, for quantitatively assessing systematic errors and other sources of uncertainty in study results.[3–9] In the form we consider here, an uncertainty analysis quantifies the impact on estimates of parameters that govern the degree of bias in the estimates, by sampling values for these parameters from prior probability distributions. Other names for this approach include Monte-Carlo uncertainty analysis, Monte-Carlo risk analysis and Monte-Carlo sensitivity analysis.

Each uncertainty analysis is situation specific, although the following four steps are common. First, specify the target parameter of interest and an equation for its conventional estimator. Second, specify an equation (or system of equations) that describes how the observed data are generated from the population of interest, including in the equation bias parameters that determine the degree of systematic error in the observations; this step is known as bias modelling.[8 9] Combining this equation with the equation for the estimator yields an equation for removing biases from the estimator. The values of the bias parameters are usually unknown; thus, as a third step, we specify prior probability distributions for the values of the bias parameters. These prior distributions describe the investigator's judgements or bets about the magnitude of the bias parameters before seeing the data.[10] These judgements should be based on careful consideration of how systematic errors might have occurred during the process of the study. The distributions are then used in a fourth step, in which we repeatedly sample from the prior distributions to produce adjusted estimates of the target parameter. This repeated-sampling (Monte-Carlo) step propagates uncertainties about the bias parameters (as expressed in the prior distributions for the bias parameters) through the estimating equations. An uncertainty analysis results in a distribution of estimates for the target parameter that accounts for the investigator's judgements about the likely magnitude of systematic errors, as well as for uncertainty about random errors.

An approach more basic than ours is sensitivity analysis.[9 11] A standard sensitivity analysis is non-probabilistic; it examines results obtained from a relatively small number of fixed values for each bias parameter. The result from a sensitivity analysis is thus a set of estimates that have been adjusted for bias under the fixed values for the bias parameters. The uncertainty analysis we describe here replaces these fixed values with many values drawn from a prior probability distribution, to represent more flexibly our uncertainty about the exact size of the bias parameter. It has been called Monte-Carlo sensitivity analysis to reflect the fact that it is equivalent to sensitivity analysis based on random draws from the prior distribution.[4 8]

A potentially more accurate approach is Bayesian uncertainty analysis.[2 4 8 12 13] This approach also uses prior distributions such as those we describe here. Instead of sampling from these distributions and using the sampled values to adjust the data or the estimates, however, Bayesian analysis updates the distributions based on the data, using Bayes' theorem, and samples the parameters from these updated (posterior) distributions. The result is a posterior probability distribution for the parameters of interest. Bayesian analysis has the advantage of being able to incorporate prior distributions for parameters other than bias parameters, including a prior distribution for the target parameter. Nonetheless, if these other prior distributions are not included in the Bayesian analysis, it often tends to give results similar to Monte-Carlo uncertainty analysis,[4 8 12] and so we do not discuss it further here.

The present article provides a teaching example of Monte-Carlo uncertainty analysis, using a simple problem of estimating four proportions from a survey of the epidemiological literature.

METHODS

Survey example

We conducted a literature survey to determine how epidemiologists assess exposure-measurement error (EME) in study results.[14] We randomly sampled 57 articles from *Epidemiology* (2001), *American Journal of Epidemiology* (2001) and *International Journal of Epidemiology* (December 2000–October 2001). Each article was reviewed by one of us (AMJ) for (1) acknowledgement of the possibility of error in measuring any of the

....

Abbreviations: CL, confidence limit; EME, exposure-measurement error; FN, false negative; FP, false positive

study exposures and (2) qualitative or quantitative evaluations of the impact of EME on the study results.

Of the 57 articles that were examined, 35 articles (61%, 95% confidence interval (CI) 49, 74) acknowledged EME. Eighteen of these 35 articles (32%, 95% CI 20, 44) mentioned EME but said nothing about the impact of EME on the study results; another 16 articles (28%, 95% CI 16, 40) qualitatively described the effect EME could have had on the study results. One study (2%, 95% CI 0, 9) quantified the impact of EME on the study results by using a sensitivity analysis.

We present the results of our uncertainty analyses, which quantitatively examined the bias in our survey results that could have been caused by errors in classifying the articles. The goal of our analyses was to obtain interval estimates for the actual proportions that accounted for both systematic error (due to the possibility that our classification of the articles was imperfect) and random-sampling error (due to the fact that we randomly sampled 57 articles from a much larger set of articles that met our inclusion criteria).

Step 1: defining the target parameters and their estimators

For the purpose of this illustration, we will define the target parameters as the proportions of articles in the surveyed journals and time period that

(1) mentioned anything about EME for their study exposures (category 1, acknowledge EME),
(2) mentioned EME but said nothing about the impact of EME on study results (category 2, no impact of EME),
(3) qualitatively evaluated the effect of EME on study results (category 3, qualitative), or
(4) quantified the impact of EME on study results (category 4, quantitative).

We denote these four proportions by P_i, where $i = 1, \ldots, 4$ denotes one of the above categories. Because we sampled 57 articles, the conventional estimator for each of these four proportions (ie, the estimator obtained by assuming no bias) is given by

$$\frac{\text{number of articles classified as in category } i}{57}. \tag{1}$$

Step 2: defining the data-generation process

For the random-error component, we assumed that the number of sampled articles falling into category i followed a binomial distribution with probability P_i and total trials equal to 57 (thus neglecting any finite-population correction, which would have slightly narrowed the CIs). For the systematic component of error, the number of articles classified as falling into a category is equal to the true number in that category minus false positives (FPs, those that were not in the category but mistakenly placed there) plus false negatives (FNs, those that were in the category but mistakenly left out). That is

$$\begin{aligned}\text{number of articles actually in category } i = \\ \text{number of articles classified into caregory } i - \text{FP}_i + \text{FN}_i. \tag{2}\end{aligned}$$

In this context, FP_i and FN_i are the bias parameters. We are not certain that our classification procedure was perfect; we therefore did not want to assume that the number of FPs was always zero, nor that the number of FNs was always zero.

We incorporated equation (2) into expression (1) to get the following four equations (one each for $i = 1, 2, 3, 4$) to estimate

the proportions while adjusting for possible classification errors:

$$\hat{P}_i = \frac{\text{Number of articles we classified into category } i - \text{FP}_i + \text{FN}_i}{57}. \tag{3}$$

Step 3: specifying prior distributions

For the classification errors that we believe could have occurred (table 1), we are somewhat uncertain about the actual numbers of misclassified articles. We specified prior distributions for these classification errors (table 1), which describe our assumptions about the likelihood and magnitude of these errors. Standard probability distributions did not adequately model our uncertainty. Instead, we specified our priors as discrete distributions, customised to match our beliefs and background information. We think it plausible that our examination of the surveyed articles resulted in little misclassification. Therefore, for each prior distribution we assigned a high probability to having zero articles misclassified.

We assumed that no more than three articles could have been classified incorrectly as ignoring EME (error FN_1; table 1); we thought that our careful reading of the articles would make a larger number highly unlikely. Given this constraint, we further assumed only one article that truly acknowledged EME with no qualitative evaluation could have been misclassified as ignoring EME (error FN_2); we thought it would have been difficult for us to miss a statement of error in physical measurements or a limitation about exposure measurement. Finally, we thought it difficult to incorrectly classify articles with qualitative statements about the impact of EME on study results either as not qualitatively evaluating EME (errors FP_2 and FN_{3B}) or as not acknowledging EME (error FN_{3A}; table 1), although unclear wording in an article might have caused these errors.

We judged the following types of classification errors to be extremely unlikely and, therefore, assumed in our analysis that they did not occur:

(a) an article that said nothing about EME was incorrectly classified as saying something about EME, or qualitatively or quantitatively evaluating the impact of EME on study results;
(b) an article that mentioned EME, but did not evaluate the impact of EME on study results was incorrectly classified as qualitatively or quantitatively evaluating the impact of EME on study results;
(c) an article that qualitatively evaluated the impact of EME on study results was incorrectly classified as quantitatively evaluating the impact of EME on study results; and
(d) an article that quantitatively evaluated the impact of EME on study results was incorrectly classified into any other category.

We believe errors (a), (b) and (c) to be unlikely because they could have occurred only if we imagined reading a statement that was not there. Error (d) is unlikely as the discussion of a quantitative evaluation would take up at least several sentences in a manuscript, or most likely even more space. Note that omitting a bias parameter assigns a prior probability of one to the null value for that parameter.

Assumptions a–d lead to the following modifications of the estimating equations (3) for the four target parameters:

$$\hat{P}_1 = \frac{35 + \text{FN}_1}{57},$$

Table 1 Descriptions of the prior probability distributions used for possible classification errors

Error	Description of error	Number of articles misclassified	Probability of number misclassified
FN_1	Number of articles that acknowledged EME, but were misclassified as ignoring EME	0 1 2 3	0.600 0.300 0.075 0.025
FN_2	Number of articles that acknowledged EME and did not evaluate its impact on study results, but were misclassified as ignoring EME	0 1	0.700 0.300
FP_2	Number of articles that qualitatively evaluated the impact of EME on study results, but were misclassified as acknowledging EME with no evaluation	0 1 2	0.800 0.150 0.050
FN_{3A}	Number of articles that qualitatively evaluated the impact of EME on study results, but were misclassified as ignoring EME	0 1 2	0.700 0.250 0.050
FN_{3B}	Same as FP_2		

EME, exposure-measurement error; FN, false negative; FP, false positive.

$$\hat{P}_2 = \frac{18 - FP_2 + FN_2}{57},$$

$$\hat{P}_3 = \frac{16 + FN_{3A} + FN_{3B}}{57}, \text{ and}$$

$$\hat{P}_4 = 1/57.$$

Step 4: Monte-Carlo uncertainty analysis

The software package Crystal Ball (version 2000.2) was used to draw samples from the distributions given in table 1.[15] Each of the four analyses was simulated separately and was based on 100 000 resamplings from these distributions.

To account for random-sampling error in our results (due to the fact that we randomly sampled 57 articles from a very large set of articles that would meet the inclusion criteria), we added a bootstrap step to each of the Monte-Carlo trials. The bootstrap step was implemented on each simulation trial by generating one binomial (57, P) random variate, where P is the corrected proportion for that Monte-Carlo trial, and by replacing the corrected number of articles with the binomial random variate.[16] To complete the bootstrap step, we corrected our estimator to reduce its variance by subtracting off excess random error. Thus, the analysis for each of the four proportions generated a distribution of 100 000 proportions accounting for classification error (based on our priors about the magnitude of classification errors) and random-sampling error.

We took the lower 2.5 and upper 97.5 percentiles of the total simulation distribution as defining our 95% simulation interval for the proportion. As mentioned earlier, in some circumstances the interval between these limits approximates a 95% Bayesian posterior probability interval—that is, we would bet 95% that the true proportion of the sampled population falls within the interval.[8] When this approximation holds, the interval is also known as a 95% certainty interval or 95% credibility interval, and can be said to fairly represent the certainty that we should have about the true proportion in light of our data and our prior distributions.

RESULTS

Table 2 presents the conventional point estimates (which are the observed proportions $\hat{P}_1$, $\hat{P}_2$ and $\hat{P}_3$) and 95% CIs for the

actual proportions of articles published in categories 1, 2 and 3, alongside the simulation medians and 95% simulation intervals for comparison (fig 1).

Our prior assumption that at most three articles were misclassified per group limited the amount of systematic error. There is a very small difference between $\hat{P}_1$ and the median of the simulation distribution. For $\hat{P}_2$ and $\hat{P}_3$, there is no discernible difference between the conventional estimates and corresponding simulation medians. Similarly, there was little difference between the 95% confidence and simulation intervals. Therefore, with our priors, sampling error is the largest contributor to total uncertainty.

DISCUSSION

We recommend using quantitative methods such as uncertainty analysis to evaluate the effect of systematic error on study results whenever policy decisions are to be based on the results. p Values and 95% CIs provide a sense of the potential impact of
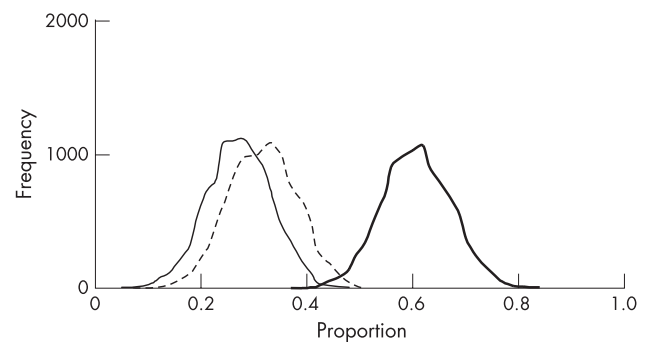
Acknowledged exposure-measurement error—uncertainty distribution for $\hat{P}_1$ = 0.61

Acknowledged exposure-measurement error but provided no evaluation—uncertainty distribution for $\hat{P}_2$ = 0.32

Qualitatively evaluated the impact of exposure-measurement error on study results—uncertainty distribution for $\hat{P}_3$ = 0.28

Figure 1 Frequency distribution representing the uncertainty in the proportion of articles that said something about exposure-measurement error by the type of evaluation.

Table 2 CIs and simulation medians and intervals for observed proportions 1, 2 and 3

Proportion number	Observed proportion	Median for simulation distribution	95% CI	95% simulation interval
1	0.61	0.63	(0.49 to 0.74)	(0.47, 0.74)
2	0.32	0.32	(0.20 to 0.44)	(0.19, 0.44)
3	0.28	0.28	(0.16 to 0.40)	(0.14, 0.39)

random error on estimates. Nonetheless, the justification for the assumptions used to compute p values and 95% CIs is not always clear,[8 17] and they do not account for the impact of uncontrolled systematic errors on estimates. Here we have illustrated an uncertainty analysis for each of four proportions by using a basic example involving one form of systematic error: misclassification. This example is much simpler than evaluating an effect measure in an epidemiological study. We think that it is valuable to start with a very simple case study of one source of systematic error, free of causal-inference concerns. One may then approach the more complex effect-estimation examples that have appeared in the literature.[2–8 12 13] Perhaps the simplest of these uncertainty-analysis examples involves a single standardised mortality ratio and a single systematic error (confounding), which we would recommend as a second teaching example.[12]

An uncertainty analysis uses prior probability distributions for the bias parameters. When many bias parameters are present, there will be a very large number of possible parameter combinations, making simple sensitivity analysis impractical or misleading. Use of prior probability distributions allow one to summarise over parameter combinations based on explicit judgements about the likely sizes of the parameters. The final simulation distribution for the target parameter is derived from the usual random error model and the assumed prior distributions for the bias parameters. Histograms illustrate how the distribution changes when one changes assumptions and prior distributions for the bias parameters.

As seen above, the most difficult step is specifying the prior distributions. Possible sources of information for the priors include eliciting expert opinion (purely subjective priors), gathering estimates from external (prior) validity and reliability studies or a combination of these sources. If no prior study data are available, the entire prior distribution will depend on the investigator's judgement. By contrast, when using a standardised instrument, reliability or validity data may be available from previous studies, and that data can be used to estimate the bias parameters (eg, false negative and false positive probabilities). Nonetheless, when previous data are available, their relevance to the current study will be a matter of judgement. Even if the data come from the current study (as in validation sampling), those data may be subject to considerable systematic error, such as selection bias, and an uncertainty analysis of these sources of bias may be warranted.

Investigators will rarely have identical priors, for they will have differing opinions about the likely sizes of bias parameters and the relevance of previous data to the current study. Thus, the results of each investigator's uncertainty analysis could conflict. This is no different from the usual disputes over study results; the advantage of uncertainty analysis, however, is that it enables one to pinpoint the major sources of conflict.

Performing sensitivity analyses can help epidemiologists explore the potential impact of systematic errors on their effect estimates. By incorporating prior distributions for the bias parameters into their analyses, epidemiologists can further study the impact of judgements about the sizes of these parameters. By varying these distributions to match different beliefs, they may also discover which (if any) conclusions are warranted in light of current beliefs about these parameters, even if there is conflict among those beliefs.

What this paper adds

- This paper provides a primer for epidemiologists who are interested in accounting for bias in the results of epidemiological studies.

Policy implications

- Policy makers can use this method to help interpret the results of epidemiological studies by quantitatively accounting for biases in the study results.

Authors' affiliations
Anne M Jurek, Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota, USA
George Maldonado, Timothy R Church, Division of Environmental Health Sciences, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA
Sander Greenland, Department of Epidemiology and Department of Statistics, University of California, Los Angeles, California, USA

REFERENCES
1 **Morgan MG**, Henrion M. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. New York: Cambridge University Press, 1990:1–256.
2 **Eddy DM**, Hasselblad V, Shachter R. *Meta-analysis by the confidence profile method*. New York: Academic Press, 1992.
3 **Phillips C**, Maldonado G. Using Monte Carlo methods to quantify the multiple sources of error in studies [abstract]. *Am J Epidemiol* 1999;**149**:S17.
4 **Greenland S**. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc* 2003;**98**:47–54.
5 **Lash TL**, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational epidemiologic data. *Epidemiology* 2003;**14**:451–8.

6 **Phillips CV**. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003;**14**:459–66.
7 **Greenland S**. Interval estimation by simulation as an alternative to and extension of confidence intervals. *Int J Epidemiol* 2004;**33**:1389–97.
8 **Greenland S**. Multiple-bias modelling for analysis of observational data [with discussion]. *J R Stat Soc A* 2005;**168**:267–308.
9 **Greenland S**, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern epidemiology.* 3rd edn, Chapter 19. Philadelphia, PA: Lippincott-Raven, 2008.
10 **Greenland S**. Bayesian perspectives for epidemiological research: I. Foundations and basic methods [with comment and reply]. *Int J Epidemiol* 2006;**35**:765–78.
11 **Maldonado G**, Delzell E, Tyl S, *et al.* Occupational exposure to glycol ethers and human congenital malformations. *Int Arch Occup Environ Health* 2003;**76**:405–23.

12 **Steenland K**, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol* 2004;**160**:384–92.
13 **Greenland S**, Kheifets L. Leukemia attributable to residential magnetic fields: results from analyses allowing for study biases. *Risk Anal* 2006;**26**:471–82.
14 **Jurek AM**, Maldonado G, Greenland S, *et al.* Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol* 2006;**21**:871–6.
15 **Decisioneering**. *Crystal Ball 2000.2 user manual.* Denver, CO: Decisioneering, 2001.
16 **Vose D**. *Risk analysis: a quantitative guide*, 2nd edn. New York: John Wiley & Sons, 2000:182–91.
17 **Greenland S**. Randomization, statistics, and causal inference. *Epidemiology* 1990;**1**:421–9.